Claims

1. A method for automatically producing an optimal summary of a linear media source, comprising:

(a) parameterizing the linear media source to produce a parameterized media signal;

(b) creating a similarity array comprised of a plurality of array elements, wherein each array element includes the value of a similarity measurement between a first portion of the parameterized media signal and a second portion of the parameterized media signal;

(c) optimizing the value of a segment fitness function over the similarity array in order to find an optimal segment, wherein the segment fitness function is adapted to measure the similarity between a segment of the parameterized media signal and the entire parameterized media signal and is a mathematical function of at least a location of the segment; and

(d) selecting a portion of the linear media source as the optimal summary, the portion of the linear media source corresponding to the optimal segment.

2. The method of Claim 1, wherein Step (a) comprises:

separating the linear media source into a set of media frames;

applying a parameterization to each media frame in order to produce a feature vector representing each media frame; and

aggregating the feature vectors in order to produce the parameterized media signal.

3. The method of Claim 2, wherein:

the parameterization of a media frame includes a frequency domain transform.

4. The method of Claim 2, wherein:

the parameterization includes assigning a token value to a portion of the media frame.

5. The method of Claim 1, wherein Step (a) comprises:

reading a linear media data file containing the linear media source divided into a plurality of media frames, each media frame containing parameterized linear media information;

creating a feature vector for each media frame from the parameterized linear media information contained in each frame; and

aggregating the feature vectors in order to produce the parameterized media signal.

6. The method of Claim 1, wherein:

the similarity measurement comprises a measurement of vector similarity between a first feature vector corresponding to the first portion of the parameterized media signal and a second feature vector corresponding to the second portion of parameterized media signal.

7. The method of Claim 6, wherein:

the measurement of vector similarity comprises the Euclidean distance between feature vectors in parameter space.

8. The method of Claim 6, wherein:

the measurement of vector similarity includes the scalar (dot) product of the feature vectors.

9. The method of Claim 6, wherein:

the measurement of vector similarity includes the cosine of the angle between the feature vectors.

10. The method of Claim 6, wherein:

the measurement of vector similarity includes applying a Term-Frequency/Inverse Document Frequency weighting to the feature vectors.

11. The method of Claim 1, wherein:

the similarity measurement comprises a vector correlation of a first plurality of feature vectors corresponding to the first portion of the parameterized media signal and a second plurality of feature vectors corresponding to the second portion of the parameterized media signal.

12. The method of Claim 1, wherein:

the similarity array comprises a two dimensional matrix with each row and each column of the matrix corresponding to a portion of the

parameterized media signal, such that each matrix element includes the value of a similarity measurement between a first feature vector, the first feature vector corresponding to the portion of the parameterized media signal associated with the row of the matrix element, and a second feature vector, the second feature vector corresponding to the portion of the parameterized media signal associated with the column of the matrix element.

13. The method of Claim 1, wherein:

the segment fitness function comprises the average similarity measurement of a portion of the similarity array, the portion of the similarity array containing a plurality of similarity measurements between a candidate segment and the entire parameterized media signal.

14. The method of Claim 13, wherein:

the segment fitness function further comprises a weighting function which emphasizes the similarity measurement for at least one portion of the parameterized media signal corresponding to a desirable portion of the linear media source.

15. The method of Claim 1, wherein Step (c) comprises:

optimizing the value of a segment fitness function using a one-dimensional optimization to find an optimal segment location for a segment of a predetermined length.

16. The method of Claim 1, wherein Step (c) comprises:

optimizing the value of a segment fitness function using a one-dimensional optimization to find an optimal segment length for a segment of a predetermined location.

17. The method of Claim 1, wherein Step (c) comprises:

optimizing the value of a segment fitness function using a two-dimensional optimization to find an optimal segment location and an optimal segment length.

18. The method of Claim 1, further comprising the steps of:

(e) removing the optimal segment from the similarity array to produce a modified similarity array;

(f) repeating Steps (b), (c), and (d) with the modified similarity array to produce a second optimal summary of the linear media source.

19. The method of Claim 1, wherein:

the linear media source includes video.

20. The method of Claim 1, wherein:

the linear media source includes audio.

21. The method of Claim 1, wherein:

the linear media source includes text information.

22. An article of manufacture including an information storage medium wherein is stored information, the information comprising:

a group of processor readable instructions adapted to operate on a processing device, wherein the group of processor readable instructions are adapted to operate the processing device according to the method of Claim 1.